ABSTRACT OF THE DISCLOSURE

Systems and methods create high quality audio-centric, image-centric, and integrated audio-visual summaries by seamlessly integrating image, audio, and text features extracted from input video. Integrated summarization may be employed when strict synchronization of audio and image content is not required. Video programming which requires synchronization of the audio content and the image content may be summarized using either an audio-centric or an image-centric approach. Both a machine learning-based approach and an alternative, heuristics-based approach are disclosed. Numerous probabilistic methods may be employed with the machine learning-based learning approach, such as naïve Bayes, decision tree, neural networks, and maximum entropy. To create an integrated audio-visual summary using the alternative, heuristics-based approach, a maximum-bipartite-matching approach is disclosed by way of example.